SEARCH ENGINES

Internet Search Engines and Strategies: A Review

Elizabeth B. Stroup
Web Librarian, Digital Library
Naval Research Laboratory
Washington, DC
October 1999

Less than 35% of the Web is currently indexed by any 1 search engine

No one web tool catalogs or organizes the whole web. You are searching and viewing data extracted from the web which has been placed into a

DATABASE

The most common problem in searching results from

SEARCHER ERROR

Searcher Errors

Common Searcher Mistakes

- Incorrect spelling or typo
- Poorly described or conceptualized topic
- Query is too general or too specific
- Search syntax is wrong or does not perform the way you expect it to perform

Web Directories and Search Engines

Directories are compiled by people

Search engines are automated

Web Directories

Directories

A directory is any listing of links compiled by SOMEONE

Directories

Most directories are browsable.

Directories usually use controlled language

Directories

- Sites attempting to organize a very large amount of information
- Sites that specialize in a limited subject area
- Distinguish themselves by the type of "value" they add to an otherwise unexciting listing

Directories

A web directory is a subject tree or subject directory - a listing of subjects organized into categories which are usually arranged hierarchically:

Computers
Internet
Web Development
Web Graphics

Lists may be topically or alphabetically organized, and have many different formats.

Directories

Use a directory

- When you want to see what is available on a topic
- When you are beginning your research
- When you trust the compiler of the directory to channel you to the best sites

SEARCH ENGINES

Search Engines

A search engine creates computer-compiled lists of words and images

Search Engines

Simple and Meta Engines

Simple = single engine search tool

Meta = multiple engines queried at once; a type of subject catalog

Search Engines

- Search engines index words and images using “spiders”, “bots”, or “crawlers”
- Search engines are powerful tools for finding information not considered by a human to be the main "topic" of a web page
- Search engines generally do not have bias
- The algorithm each search engine uses to bring back results is different from all others and proprietary to itself

Search Engines

Search the database of a search engine by entering a key word in its dialog box.

Results are displayed as a list of web sites in which these terms occur; they are presented as relevant documents -

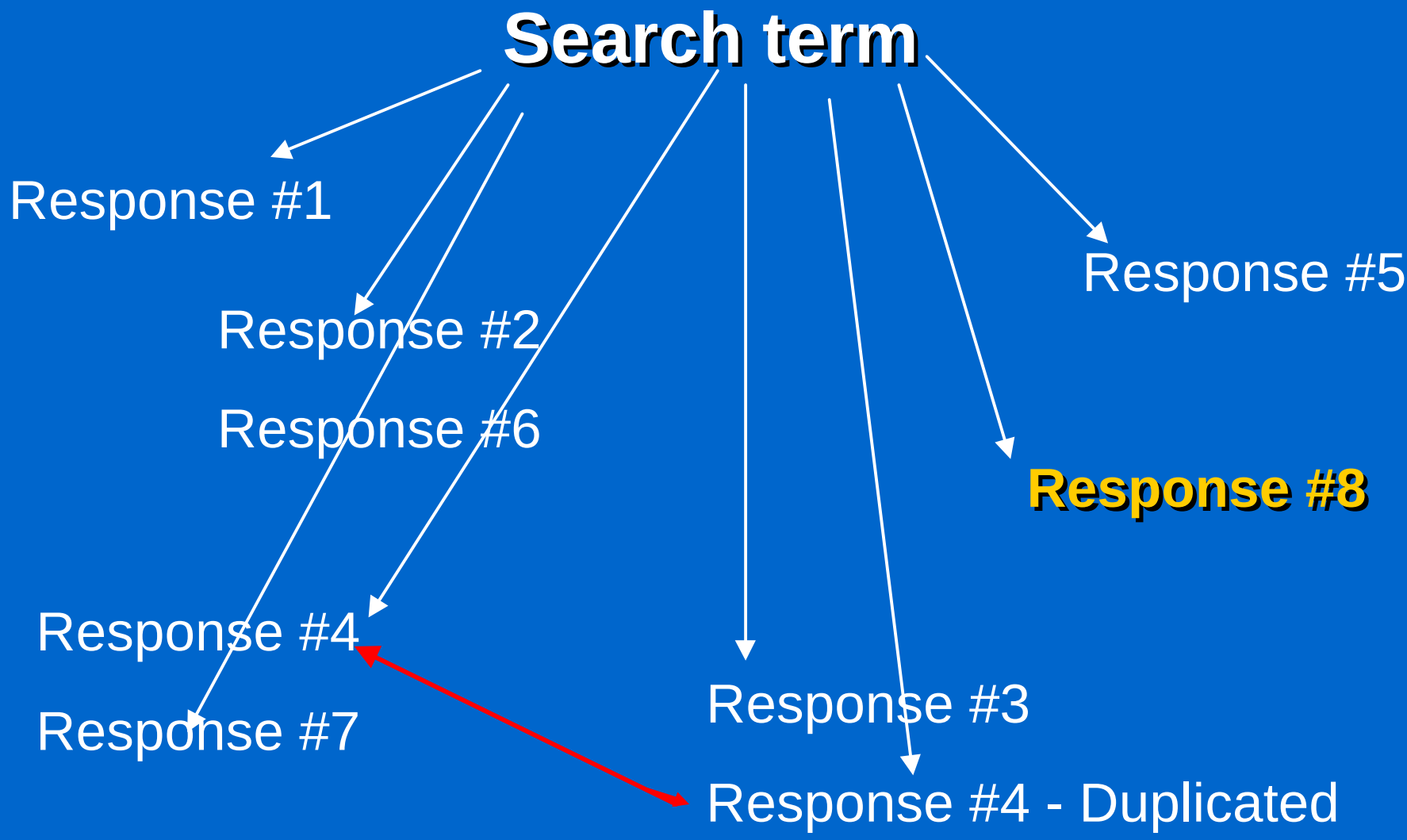
regardless of context

Search Engines

Although some search engines use ranking, most search engines present their results

IN RANDOM ORDER

Search Engines



The response you wanted is #8

Search Engines

Things to look for in a search engine:

- Visible “HELP” information
- Customization options available
- Elimination of dead links from retrieval results
- Elimination of duplicate links from retrieval results

Search Engines

Use a search engine

- When time is not a factor
- When you'd like to see new sites
- To continue your research
- With a keyword search to access material that may not otherwise be “found”

Search Engines

More is not always better

Search Engines

Things to look for in a meta search engine

- A display of search engines queried
- Customization options/choice
- Saving the result set
- Topical, subject, specialty links added to searches
- Query features

Search Engines

Query Features

- Interpret the search string into the native language of the engine searched
- Truncation - * / % / auto plurals
- Case sensitivity
- Proximity features

Key Word Searching

Precision is King

Keyword search for precision

Boolean operators can aid this process if allowed by the engine

Boolean operators include:

and / or / not / + / - / ( ) / “ “ / and not

Key Word Searching

AND - retrieves information containing both terms/phrases

OR - retrieves information with either term/phrase

NOT - eliminates term/phrase following it

+ indicates a term that must be included in results

- indicates a term that is not important to results

( ) or " " indicate a wholly-inclusive search string

AND NOT - indicates a subgroup to be eliminated from results

Key Word Searching

Use the words or phrases that best fit your topic

- If you're looking for a **bathysphere**, use that term;
  do not type "diving bell"
- If you want information specific to the
  Lake Tahoe Bathysphere
  include all those words in your initial search string; use engine-allowed search string or important word indicators

Adding words to a search string usually increases the number of items retrieved without these

Key Word Searching

- If you don’t get the information you want, try different key words
- If you need help, use the “HELP”, “FAQ”, “ABOUT”, “HOW TO SEARCH”, or other guides provided by the search engine itself for search tips and use
- Check your spelling and punctuation
- If one search engine does not work, there are others: **TRY THEM!**

Key Word Searching

The Key Word Queen is fickle

Advantages and Disadvantages of key word searches

- Best used when what you want is unusual, unique, specific, or obscure
- May result in multiple same-site referrals
- May overwhelm the user with references that are totally irrelevant to the search and require a lot of time to sort through.

Search Engines

SIMPLE ENGINES

Single Engines

Excite

Can be good but is limited in what it can do an improved version is in the works

Does not handle complex searches well

Boolean - and, or, not, (), +, -

Truncation - no

Case sensitive - no

Single Engines

FAST

Searches over 2 million pages and 30% of the web; currently the largest simple engine

Best used for simple, focused searches

Boolean = +, -

Truncation - no

Case sensitive - no

Single Engines

Hot Bot

Handles most searches well, even detailed queries

Clusters results by domain

Is case-sensitive

Boolean = and, or, not, (), +, -

Truncation - *

Single Engines

Jayde Science

Directory of various science-related areas

Uses simple, proximity, phrase, “and”, “or” search methods

Boolean = and, or

Allows some customization through Pro Search option

Contains 20 main web guides and 200+ sub-categories of “mainstream interest”

Single Engines

Lycos

Good simple searches but may require tweeking for best results

Difficult to access in pre-peak as well as peak times

Boolean = and, or, not, (), +, -

Truncation - no

Case sensitive - no

Single Engines

Google

Simple engine good for general and complex searches

Does not always handle narrow searches well

Clusters results

Boolean +, - no truncation or alternate searches

Check Google ScoutLink for pages like the best response

Read “about” for find list of STOP words

Case sensitive - no

Single Engines

Northern Light

Second largest simple search engine

Free and fee-based results

Clusters results in subject folders for ease of location

Boolean = and, or, not, (), +, -

Truncation - *, %, auto plurals

Case sensitive - no

Single Engines

Snap

Human-compiled directory-type engine

Browse for broad subject areas, then search

Uses key word and boolean operator options for refining searches

Boolean = and, or, not

Truncation - *

Case sensitive - yes

Single Engines

WebCrawler

Simple engine search

Interprets search queries literally

Boolean = and, or, not, (), +, -

Truncation - no

Case sensitive - no

META ENGINES

Meta Engines

Ask Jeeves!

The Internet “Butler”

Use natural language - type the question

Queries own knowledgebase and other engines for results

Eliminates duplicate responses from engines searched

Works best with uncomplicated searches

Use boolean “ “ to show words that belong together in string

Meta Engines

Dogpile

Searches 9 engines

User sets order of engines you search - search 3 at a time,
then moving to next 3, and last 3 as you request

Key word searches may use “ “ or ( ) to indicate word groupings

Does not eliminate same-site response duplications

MetaFind feature

Meta Engines

GoCrawl Science Search

Directory of subject areas w/ subheadings

Keyword searching of 100s of science sites

Truncation *

Minimal customizing in results format

Meta Engines

Mamma

Searches 10 engines simultaneously

Reformats queries in appropriate syntax for each engine

“Power Search” option
- List of engines for selection
- Phrase selection option
- Search Tips
  - Document title searching

Difficult to reach during peak usage times

Meta Engines

MetaCrawler

- Searches top Web engines
- Eliminates duplication, scores results, provides list of relevant sites
- Power search & customization options
- “ “ option for word strings
- Boolean = +, -

Meta Engines

ProFusion

Searches 9 engines

You select search options - all; best 3; fastest; choose 3

Customizations features prominent

Boolean operators in advanced search must be UPPERCASE

Phrases supported in all engines ProFusion queries

Meta Engines

Savvy Search

- Searches up to 100 engines - not a good idea!
- Customize by engine order and number; also topic or subject
- Search access to guides and news databases and specialty areas
- Allows user to save results for future use
- Contains 20+ foreign language mirror sites

Meta Engines

Highway61

Searches 5 engines

Boolean AND OR +

Hits “Lots” = up to 70, “Bury me” =60-125

Ranking based on how many sites an item was found at

Eliminates duplication in results

Saves search parameters (and patience level) as a cookie for future use

Cookies

DO NOT DISABLE THE COOKIES FEATURE!

If you do this, you will be denied access to NRL-licensed publishers' Web sites. IP-recognition, necessary to the access process for NRL users, does not work when the cookies feature is disabled.

Cookies are harmless. With newer versions of Internet browsers, placement and receipt of a cookie is not obvious. Cookies are placed on a computer's hard drive by Web site servers for use as tracking mechanisms. If their presence bothers you, ask your computer support personnel to delete them for you.